ChinaXiv合作期刊
图书情报工作
第66卷 第2期 2022年1月

大数据专业培养内容的主题分析及对图情档学科的启示*

■ 杨杰[1] 赵星[1,2]

[1] 华东师范大学经济与管理学部信息管理系 上海 200062 [2]华东师范大学学术评价与促进研究中心 上海 200062

摘 要: [目的/意义]在大数据浪潮和"新文科"背景下,中国图情档学科的人才培养范式亟需改革。与此同时,大数据相关专业的建设方兴未艾,对于图情档学科的人才培养新范式建设具有借鉴意义。[方法/过程]采用一种时序主题网络模型及计算方法;通过搜集、处理、统计、分析259所高等院校的大数据专业培养方案文本,在时间维度上进行主题挖掘,总结归纳数据科学课程的层次;分析图情档学科主干知识与大数据专业的联系,并给出适合图情档学科的数据科学课程建议。[结果/结论]结果表明,所采用的时序主题网络模型方法能够较好地契合大数据专业的人才培养范式发展分析,可以成为研究学科主题的一种方法。此外,还给出面向图情档学科的数据科学类课程建议,可供图情档学科的人才培养参考。

关键词: 人才培养 数据科学 新文科 时序主题网络模型

分类号: G254.9

DOI: 10.13266/j.issn.0252-3116.2022.02.012

1 引言

"新文科"建设是提升中国国家软实力、促进文化繁荣的重要部分,也是高质量教育体系改革的重点。育人育才是建设"新文科"的关键点,要做到遵循规律、协同育人、模式多元,培养出适应新时代要求的应用型、复合型人才。

图书情报与档案管理(以下简称"图情档")学科因为具有天然的文理交叉性,在"新文科"的背景下,人才培养模式必然迎来重大的革新。在数据科学刚刚兴起时,叶鹰和马费成指出数据科学与信息科学在理论逻辑和技术方法上一脉相承[1]。"数智赋能"已然成为图情档学科的新发展趋势,孙建军等[2]认为数据管理与数据分析技术为图情档学科的发展提供了新的可能,赵星等[3]也基于数据智能和知识发现提出了图情档学科的拓展方向。

本研究从图情档学科发展难点出发,借鉴大数据专业人才培养模式的建设经验。为分析出大数据专业人才培养发展范式,本研究创新性地利用时序主题网络模型,对259所开设大数据专业院校的培养方案文本进行挖掘,并总结出其课程的三大层次。为助力图情档学科人才培养模式的创新建设,本研究结合叶鹰[4]总结的图情学科主干知识,探索图情档的数据科学类课程建设方案,为图情档学科人才培养模式的创新发展提供依据。

2 困境与思考

2.1 现实问题

从大多数院校图情档专业所在院系设立的本科专业——信息管理与信息系统专业的发展现状来看,2020年、2021年分别有13所、16所院校撤销了该专业,在普通高校新增撤销本科专业名单中分别排在第3名、第2名。相比之下,大数据专业的开设情况却如火如荼,近两年申请开设数据科学与大数据技术专业、大数据管理与应用专业的学校均超过50所。图情档相关专业最悲观的预测是在10年内"消亡、被合并、被替代、改名、被边缘化"[5]。由此可见,当前图情档学科的人才培养模式亟需改革,要顺应"新文科"的时代洪流推进发展。

2.2 学界思考

自"新文科"提出以来,很多学者都对图情档学科的未来发展提出了新的思考,如初景利[6]提出图情档学科要发展成"硬学科",要做到坚持内核的情况下,强化技术与方法的引入。马费成等[7]提出要抓住新文

* 本文系国家自然科学基金面上项目"跨维度引文分析方法研究"(项目编号:71874056)研究成果之一。

作者简介: 杨杰,本科生,E-mail:alexjieyang@outlook.com;赵星,经济与管理学部副主任,学术评价与促进研究中心主任,教授。

收稿日期:2021-06-01 修回日期:2021-09-16 本文起止页码:109-116 本文责任编辑:王传清

科建设契机，重视交叉融合，同时要坚守人文传统。张九珍[8]提出图情档在“新文科”背景下的学科交叉融合应当做到“以我为主，为我所用”。关于坚持学科内核方面的探讨，叶鹰[4]对图书情报学的主干知识和研究方法进行了总结，为图情档学科的发展明确坚守的依据。周文杰[9]提到，图情档新文科建设需要坚守的“旧”内核包括：基于科学数据建设数据科学、作为基础设施支持数字人文、立足知识组织“编制”数字记忆、坚守公共文化服务的人文情怀、开掘参考咨询的循证决策功能。

2.3 创新发展

“新文科”的新是创新的新，“新文科”专业人才培养模式的建设，可以借鉴和吸取“新工科”专业的新方法、新技术。图情档学科人才培养新模式的建设，可以结合大数据专业，探索出新的发展方向。

从大数据专业汲取经验的思路并非本研究原创，陈沫等[10]研究了情报学取向的数据科学与大数据专业，依次划分出了基础课程、方法类课程和应用型课程。陶俊等[11]分析了国外多所 iSchool 院校的数据科学专业培养方案，为图书情报的数据科学课程建设提供建议。赵星等[12]利用内容分析法研究了大数据管理与应用专业的培养方案文本。黎海波等[13]研究了信管专业的数据科学课程群建设，为信管专业人才培养模式提供数据科学能力培养的新思路。闫慧等[14]通过设计民族志未来访谈，判断出数据科学与图情学科领域将来最可能是并集的关系。

本研究将从大数据专业的人才培养模式中进行探索，为图情档学科人才培养模式提供新思路。

3 大数据专业的主题分析

3.1 大数据专业的基本情况

国内大数据专业最早开设于 2016 年，具体开设情况如图 1 所示，在 2016 年，北京大学、对外经济贸易大学、中南大学首先开设了数据科学与大数据技术专业，此后在 2017 – 2021 年，分别有 32 所、248 所、203 所、143 所、62 所院校开设了该专业。该专业授予的学位分为工学学位和理学学位两种，厦门大学、中国人民大学、上海财经大学等高校同时在理学和工学对应的学院开设该专业。截至 2021 年初，开设该专业的院校数量为 674 所，其中，原“985 工程”院校有 29 所，原“211 工程”院校有 73 所。大数据管理与应用专业诞生于 2017 年，2018 年至 2021 年的开设院校数量分别为 5 所、25 所、52 所、68 所，截至 2021 年，共有 140 所高校开设了该专业，其中，原“985 工程”院校有 5 所，原“211 工程”院校有 25 所，该专业授予的学位均为管理学学位。

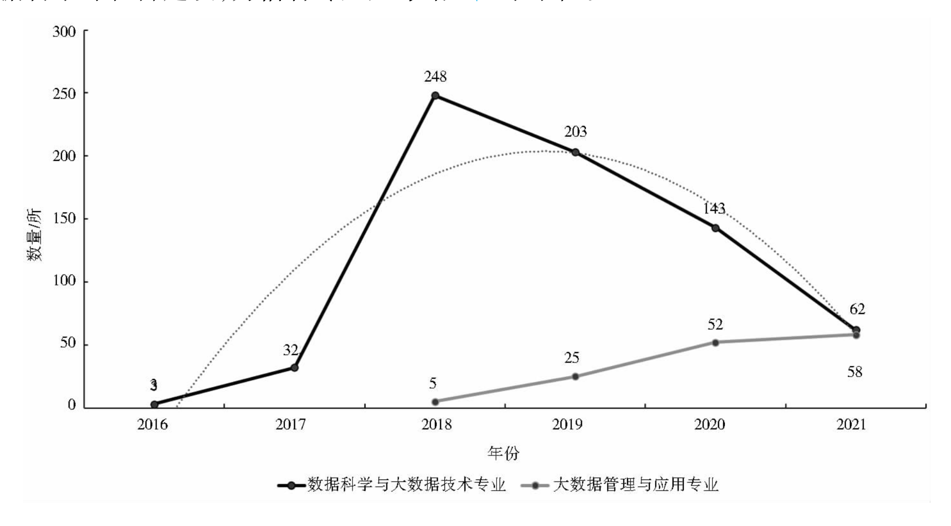


图 1 大数据专业申请开设院校数量与趋势

本研究通过邮件咨询、院校官网等渠道，对以上大数据专业开设院校的培养方案进行搜集。其中部分院校的数据难以获取，最终搜集得到了 86 所大数据管理与应用专业开设院校，173 所数据科学与大数据技术专业开设院校，共 259 所院校的大数据专业培养方案和核心课程。涵盖一流大学建设高校（如北京大学）、一流学科建设高校（如华中师范大学）、其他省部共建高校（如广东海洋大学）、省属高校（如聊城大学）、独立院校（如集美大学诚毅学院）等多个层次的高校。

3.2 大数据专业人才培养模式的主题关系

为分析大数据专业培养模式的核心主题及整体架构，本研究基于共现技术研究大数据专业培养模式各

杨杰，赵星. 大数据专业培养内容的主题分析及对图情档学科的启示[J]. 图书情报工作,2022,66(2):109－116.

个主题的之间的相互关系,将培养方案文本进行分词、去除停用词、主题标引等预处理后,对核心主题进行主题关系挖掘,得到大数据专业人才培养模式主题关系弦图,见图 2。

图 2 分别呈现了数据科学与大数据技术专业、大数据管理与应用专业的人才培养模式主题,弦的不同弧线代表该专业培养方案中的重要主题,弧线长度代表主题的重要程度,弦网络的密集程度代表不同主题之间的关联紧密性。由图 2 分析可知,数据科学与大数据技术专业人才培养的核心点为数据分析,专业核心技术是计算机技术和统计学。大数据管理与应用专业人才培养的核心点是数据管理,专业核心技术是统计学、管理科学和计算机科学。前者重视培养学生的抽象思维能力、数学形式化思维能力、数据科学理论知识、数据科学基础能力等。后者重点培养学生的数据采集、数据处理和分析、数据可视化能力。在大数据专业的培养模式下,学生能够采用数据科学方法对商业、管理中的复杂问题或复杂数据工程问题进行研究。

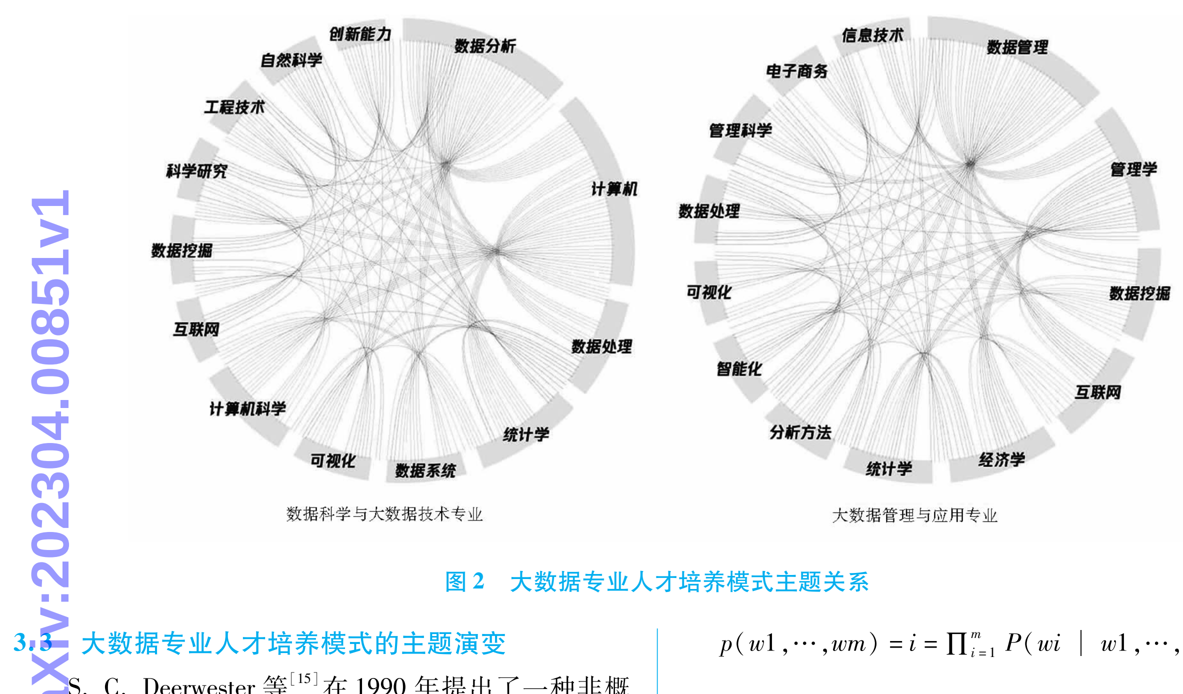


图 2　大数据专业人才培养模式主题关系

3.3　大数据专业人才培养模式的主题演变

S. C. Deerwester 等[15]在 1990 年提出了一种非概率的主题模型潜在语义索引(Latent Semantic Indexing, LSI),而后 T. Hofmann[16]实现了基于概率的主题模型,此后基于概率的主题模型逐渐兴起。传统的主题模型不包含时间维度,本研究借鉴廖君华等[17]、A. Bruns[18]、M. J. Westgate 等[19]、Z. F. Zhang 等[20]、Y. Zheng 等[21]研究的时序主题模型,利用切片方法引入时间维度。

为了从时间维度上深入剖析大数据专业的范式发展脉络,本研究利用 Python-3.8.5、Gephi-0.9.2、VOSviewer-1.6.16 软件,基于 ForceAtlas2 算法[22]计算布局,其中具体计算方法如下:

(1)根据不同时间段将文本数据集进行切片,共分为 j 个切片。

$$m_t = \sum_{i=1}^{i=j} m_i \qquad \text{公式(1)}$$

(2)对每一个切片中的文本数据进行去停用词、词频统计、词共现矩阵、网络分布,并基于概率计算主题分布。

$$p(w1, \cdots, wm) = i = \prod_{i=1}^{m} P(wi \mid w1, \cdots, wi-1)$$

公式(2)

其中,p 表示词向量编码中词组合的出现概率,将词共现矩阵的行作为词向量,并降维处理为节点数据和边数据,以方便后续可视化。

(3)对全部网络分布进行组合、叠加,得到时序主题网络。

$$TRAN_x(n) = |X(n) - X(n-1)| + \frac{x_t n}{j}$$

公式(3)

其中,$TRAN_x(n)$表示某切片数据集的共现网络在时序网络中的某一维度的位置偏移量,此偏移量由相邻两个切片数据集的共现网络的范围的差值$|X(n) - X(n-1)|$和该切片的位次 n 决定。

对大数据专业培养方案的文本数据进行如下分析:①根据院校开设大数据专业的年份对文本数据进行切片。②基于哈工大停用词表和自制词表去除停用词。③利用基于 python 的 pandas 包计算词频,利用词向量计算共现矩阵。④利用 VOSviewer 软件计算每个切片的文本数据并进行网络分布计算。⑤利用 Gephi

对全部网络分布进行布局组合，得到大数据专业的时序主题网络，见图 3。

由图 3 分析可知，大数据专业刚刚开设时，人才培养模式上倾向于数据系统、工程项目等技术性方向。而随着更多院校开设了大数据专业，该专业的人才培养模式呈现出多元化、宽口径、开放式的特征。具体而言，培养模式文本中社会科学、统计学等主题的节点密度增加，说明在人才培养方面，大数据专业逐渐将数据科学与管理学、医学、经济学等专业深度融合，在电子商务、信息管理、金融科技、智慧医疗等领域开辟新的培养模式。人才培养范式逐渐从基础理论和基本应用拓展到多元宽泛的应用方向。

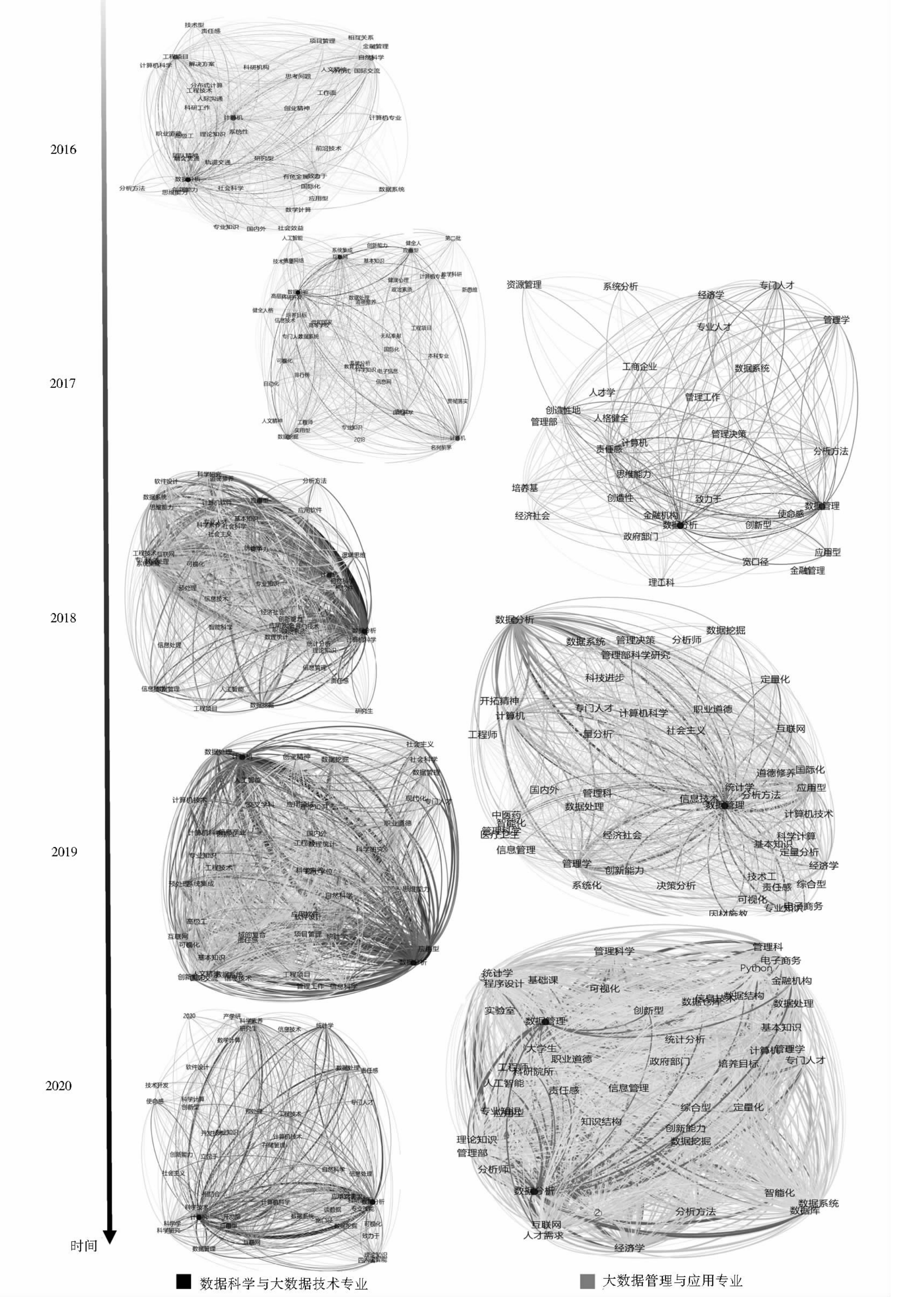


图 3 大数据专业人才培养时序主题网络

3.4 大数据专业课程的三大层次

由上文分析可知,大数据专业重视培养学生的基础数理能力、数据科学思维能力和数据科学实践能力,近年逐渐开始融合社会科学等其他学科的知识,为学生提供更加多元化的发展路径,相关课程也从基础理论知识、大数据基础应用知识演变为更加广泛的知识。

陈沫等[10]在研究大数据专业课程体系时,将其划分为基础层、方法层、应用层共三大层次。本研究借鉴其思路,并基于前两节的主题分析,将大数据专业的核心课程划分为 3 个层次,分别是基础理论与方法层、大数据理论与方法层、综合方法及应用层,见表 1。

基础理论与方法层的课程注重培养学生的基础数理思维和基本数据科学素养,包括统计学类的课程和算法程序类的课程,例如数据结构与算法、Python 程序设计与应用等课程。

大数据理论与方法层是大数据专业的核心课程,培养学生数据科学及大数据的核心素养,包括数据存储与管理、数据采集与加工、数据分析与挖掘、大数据技术 4 个方面的内容。其中,数据存储与管理包含数据库、数据仓库、Hadoop 大数据存储等课程;数据采集与加工包含爬虫、数据搜集、信息检索等课程;数据分析与挖掘包含数据建模、统计分析、机器学习、数据可视化等课程;大数据技术包含基本的大数据分析技术课程,如大数据分布式算法等。

综合方法及应用层为学生提供了多种方向的选择(不同院校根据自身情况开设),使学生运用数据科学的基本知识,在商业大数据、医学大数据等方向进行实践,能有效激发学生的兴趣,并提升学生的实践能力。例如复旦大学的大数据专业培养方案中,为学生提供了“理医工学大数据分析模块”“社会科学大数据分析模块”“类脑计算方向专业进阶模块”等多个多元发展路径的课程选择。

表 1 大数据专业课程层次

课程层次	课程类别	典型课程
基础理论与方法层	统计学原理	R 语言及统计分析、应用随机过程、时间序列分析、非参数统计、统计分析与决策
	算法与程序	C/C＋＋程序设计、Python 程序设计、操作系统原理及应用 Linux、数据结构与算法
大数据理论与方法层	数据存储与管理	Hadoop 大数据存储、Hive 数据仓库、MySqL 数据库、NoSQL 数据库、分布式数据库原理与应用
	数据采集与加工	网络爬虫与数据收集、数据采集加工与 ETL 技术、信息检索基础、智能信息检索、海量数据预处理实战
	数据分析与挖掘	数据建模与统计分析、数据挖掘原理与算法、人工智能、数据可视化技术、机器学习、自然语言处理
	大数据技术	大数据分布式算法、用户分析与推荐算法、云计算、大数据语义网络
综合方法及应用层	商业大数据	金融大数据分析、营销数据科学、商务智能方法与应用、商务智能与决策分析
	医学大数据	医疗大数据分析与应用、精准医疗统计与智慧决策
	大数据安全	信息安全、大数据安全与隐私保护、密码学原理与实践、软件安全
	其他	遥感大数据分析、旅游数据挖掘与分析、视觉与数据计算、林业大数据分析与决策

4 图情档学科人才培养的创新发展

4.1 图情档主干知识与大数据专业的联系

在汲取大数据专业建设经验,建设图情档学科人才培养新模式的过程中,要掌握图情档学科人才培养模式的核心立足点。

S. R. Ranganathan[23]很早就确立了图书馆学的立学原则,主要集中在图书服务方面。随着时代的发展,图情档学科的核心主题有了一定变化。叶鹰[4]经由核心学术凝聚表达了图情档学科的主干知识,层次上包括概念级、理论级、体系级。首先将核心学术分为信息组织、信息检索和信息分析,再分别由 3 个层次细化。信息组织在概念级细化为分类、标引,在理论级细化为分类法、主题法、编目法、索引法,在体系级细化为文献体系、知识体系。信息检索在概念级细化为查准、查全,在理论级细化为布尔检索等搜索算法,在体系级细化为搜索引擎等检索系统。信息分析在概念级细化为广、快、精、准,在理论级细化为引文分析、内容分析,在体系级细化为量化分析、质性分析。基于此,总结出图情档学科人才培养体系的核心主题,主要包括信息组织、信息检索、信息分析等方面。

在 259 所开设大数据专业院校的人才培养方案文本中分别检索、统计上述核心词,以核心词的频次 x_j 的总和作为图情档学科的核心概念出现频次总和,本研究用其来表示图情档学科和大数据专业的联系紧密程度。

$$R(n) = \sum_{i=1}^{i=j} x_j \quad \text{公式(4)}$$

分别对每个层次下的所有核心词的频次 x_j 进行求和,基于大数据专业课程层次,得到图情档主干知识及与大数据专业的联系,见表 2。

显然,数据科学已经融入了信息组织、信息检索和信息分析的各个层次,其中,“分类”“搜索算法”“量化

分析”等领域中，图情档学科的核心概念出现频次较高。由此可见，“基础理论与方法层”以及“大数据理论与方法层”的部分数据科学课程能够较好地切合图情档的核心主题，对于图情档学科人才培养新范式的建设有一定的借鉴意义。

表 2 图情档主干知识与大数据专业的联系

学科主干知识	层次	主题核心词	大数据专业培养方案中的出现频次	对应的大数据专业课程层次与类别
信息组织	概念级	信息组织	0	基础理论与方法层－算法与程序
		分类	17	大数据理论与方法层－数据分析与挖掘
		标引	2	
	理论级	分类法	3	
		主题法	1	
		编目法	0	
		索引法	0	
	体系级	文献体系	0	
		知识体系	9	
信息检索	概念级	信息检索	1	基础理论与方法层－算法与程序
		查准、查全	0	大数据理论与方法层－数据存储与管理
	理论级	布尔检索	0	大数据理论与方法层－数据采集与加工
		算法/搜索算法	18	
	体系级	搜索引擎	0	
		检索系统	0	
信息分析	概念级	信息分析	1	基础理论与方法层－算法与程序
				大数据理论与方法层－数据分析与挖掘

4.2 图情档的数据科学类课程建设

基于上一节的探讨，总结出大数据专业课程层次与图情档学科主干知识对接点，如图 4 所示。其中上方为图情档主干知识，下方为数据科学课程链，将数据科学课程链中的课程类别与相关度较高的图情档主干知识对应起来，得到相应的对接点，以此可以总结出图情档学科的数据科学类课程群。

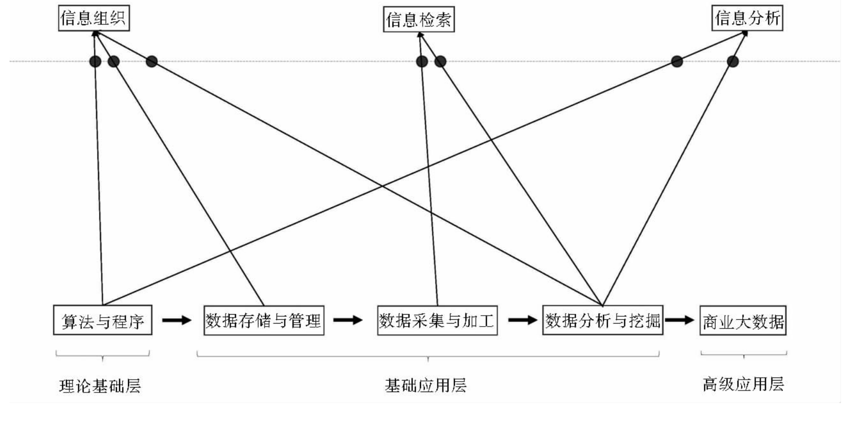


图 4 大数据专业课程层次与图情档学科主干知识对接点

面向图情档专业人才培养的数据科学类课程群，整体架构上应该更多借鉴大数据专业的课程框架，课程建设理念上要重视数据科学的应用实践。可以根据院系的具体情况，在基础理论与方法层设置 Python 程序设计；在大数据理论与方法层，也可以选择以 Python 语言为核心，依次开设数据存储与管理、数据采集与加工、数据分析与挖掘的课程，例如开设基于 Python 的网络爬虫、数据分析、数据可视化等课程。在综合方法及应用层，与商学院等其他学院的教师合作，开展商务智能等大数据课程，形成“Python 编程基础－Python 网络爬虫－Python 数据分析与可视化－商务智能”的一系列面向图情档专业的数据科学课程链。

正如柯平[24]提到的,“新文科”背景下图情档学科的人才培养模式发展不能一味地增加不合适的数据科学课程。面向图情档专业人才培养的数据科学课程群建设,要立足于信息组织、信息检索和信息分析等学科主干知识的传授。结合大数据专业的方法和技术,应用于图情档学科的原本领域,建设具有图情特色的数据科学课程群。本研究从大数据专业的培养方案中挑选出了一部分课程,可以用来培养图情档学科学生的信息组织、信息检索和信息分析能力,如表 3 所示:

表 3　面向图情档的数据科学类课程

核心培养点	课程建议
信息组织	数据结构与算法、深度学习基础、知识图谱、非结构化数据库、分类与标引、语义网络
信息检索	网络爬虫与数据收集、信息检索基础、智能信息检索、用户分析与推荐算法
信息分析	引文网络分析、数据挖掘、自然语言处理、Python 程序设计、R 语言及统计分析

需注意的是,不能简单地从大数据专业照搬一些数据科学课程,而是要结合图情档主干知识和院系的具体情况。面向图情档的数据科学课程在数据科学思维和理论层次要求不高,但要重视实践性与结合性。在培养学生信息组织能力的数据科学课程建设中,注意将知识图谱、分类与标引、语义网络等方面的内容引入相关课程;培养学生信息检索能力的数据科学课程建设中,除教学基础的数据搜集与信息检索内容外,还可以结合深度学习、强化学习等知识,在智能信息检索、用户推荐算法领域的应用加强教学;培养学生信息分析能力的数据科学课程建设中,可以设置一些算法与程序课程,例如设置基于 Jupyter Notebook 平台的 Python 程序设计教学,还可以将前沿的数据挖掘算法、自然语言处理、引文网络分析的基本内容纳入课程内容。

5　结语

本研究探讨了大数据专业的人才培养范式及主题演变网络、大数据专业的课程层次、图情档学科主干知识及与大数据专业的联系、面向图情档人才培养的数据科学类课程群建设思路以及具体的课程内容,陈沫等[10]提出的课程体系分层和叶鹰[4]提出的图情档学科主干知识对本文具有重要的基础和启发作用。

研究提到的基于数据科学建设图情档学科人才培养新模式是一种发展的路径选择,有高校已经将数据科学作为图情档学科自设的二级学科。但图情档也有不少其他的特色方向,例如中国人民大学着力建设的“数字人文”,复旦大学积极推动的“古籍保护与非物质文化遗产”及华东师范大学于 2015 年启动的“商业分析”探索等。无论发展模式如何抉择,做到守正与创新都应是改革的前提。

大数据时代下,图情档学科人才培养改革发展刚刚起步,且面临着诸多问题:如何将数据科学课程与本专业的学科特色深度结合?如何使图情档专业的人才培养模式与社会需求相结合?如何使图情档专业的人才培养模式向国际看齐,又不失中国特色?这些问题近年来已有诸多讨论,但仍未有显著共识。

此外,方法层面,本研究应用的时序主题网络模型也存在局限性,后续研究可以考虑:①完善科学的切片范围计算指标;②探索更合适的布局计算方法,以清晰、完整地呈现演变网络;③利用 PCA 降维等方法降低共现的稀疏性。

参考文献:

[1] 叶鹰,马费成. 数据科学兴起及其与信息科学的关联[J]. 情报学报,2015,34(6):575－580.
[2] 孙建军,李阳,裴雷. “数智”赋能时代图情档学科变革之思考[J]. 图书情报知识,2020(3): 22－27.
[3] 赵星,乔利利,叶鹰. 面向数据智能和知识发现的图书情报学跨界拓展——数据－学术－创造整合论[J]. 中国图书馆学报,2020,46(6):16－25.
[4] 叶鹰. 试论图书情报学的主干知识及有效方法:兼论双证法和模本法之效用[J]. 中国图书馆学报, 2021,47(3):58－66.
[5] 闫慧,韩蕾倩,吴萌,等. 图书馆学、情报学与档案学 2029 年发展前景研究[J]. 图书与情报,2019(6):2－17,153.
[6] 初景利. “新文科”呼唤图情档成为“硬”学科[J]. 图书与情报,2020(6):1－3.
[7] 马费成,李志元. 新文科背景下我国图书情报学科的发展前景[J]. 中国图书馆学报,2020,46(6):4－15.
[8] 张久珍. 图情档学科建设需要与新时代发展接轨[J]. 图书与情报,2020(6):17－18.
[9] 周文杰. 从多元异构走向融合归一——图情档新文科建设的趋向评析[J]. 情报资料工作, 2021,42(2):14－21.
[10] 陈沫,李广建,陈聪聪. 情报学取向的“数据科学与大数据技术”专业人才培养[J]. 图书情报工作,2019,63(12):5－11.
[11] 陶俊,何晓东. 面向图书情报的数据科学专业课程结构比较研究[J]. 图书馆学研究,2019(6):10－16.
[12] 赵星,俞晓婷,万玲玉. 新文科背景下大数据管理与应用专业培养特征的内容分析[J]. 图书与情报,2020(6):26－34,92.
[13] 黎海波,谢健民. 信管专业在新文科背景下的数据科学课程群建设研究[J]. 情报科学,2020,38(8):128－133.
[14] 闫慧,韩艳芳,张钰浩,等. 图书馆信息学与新文科交叉领域的关系预测研究[J]. 情报资料工作,2021,42(1):21－27.
[15] DEERWESTER S C, DUMAIS S T, LANDAUER T K, et al. Indexing by latent semantic analysis[J]. Journal of the American So-

ciety for Information Science, 1990, 41(6):391 –407.
[16] HOFMANN T. Probabilistic latent semantic indexing [C]//Proceedings of the 22nd annual international SIGIR conference. New York: ACM Press, 1999:50 –57.
[17] 廖君华,孙克迎,钟丽霞. 一种基于时序主题模型的网络热点话题演化分析系统[J]. 图书情报工作,2013,57(9):96 –102,118.
[18] BRUNS A. How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi[J]. Information, communication & society, 2012, 15(9):1323 –1351.
[19] WESTGATE M J, BARTON P S, PIERSON J C, et al. Text analysis tools for identification of emerging topics and research gaps in conservation science[J]. Conservation biology, 2015, 29(6): 1606 –1614.
[20] ZHANG Z F, LI Q D. QuestionHolic: hot topic discovery and trend analysis in community question answering systems[J]. Expert systems with applications, 2011, 38(6):6848 –6855.
[21] ZHENG Y, MENG Z P, XU C. A short-text oriented clustering method for hot topics extraction[J]. International journal of software engineering and knowledge engineering, 2015, 25(3):453 –471.
[22] JACOMY M, VENTURINI T, HEYMANN S, et al. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software[EB/OL]. [2021 –12 –09]. https://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC4051631/.
[23] RANGANATHAN S R. The five laws of library science[M]. 2nd ed. Madras: The Madras Library Association, 1957.
[24] 柯平. 新图情档——新文科建设中的图书情报与档案管理一级学科发展[J]. 情报资料工作,2021,42(1):15 –20.

作者贡献说明:
杨杰:负责数据搜集、数据分析、数据可视化、论文撰写;
赵星:负责选题拟定、论文修改。

Theme Analysis of the Training Contents of Big Data Subjects and the Enlightenment to the Library and Information Science

Yang Jie[1] Zhao Xing[1,2]
[1] Department of Information Management, East China Normal University, Shanghai 200062
[2] Institute for Academic Evaluation and Development, Shanghai 200062

Abstract: [Purpose/significance] Under the background of the big data tide and the new liberal arts, there is an urgent need to reform and innovate the talent training mode of library and information science in China. The construction of the big data subjects is in the ascendant, which has a strong reference significance for the construction of a new paradigm of talent cultivation in library and information science. [Method/process] This paper innovatively proposed a new sequential topic network model and the calculation method. By collecting, processing, counting and analyzing the talent cultivation policies of the big data subjects in 259 universities, this paper conducted topic mining in the time dimension and summarized the levels of data science courses. Additionally, this paper analyzed the relevance between the subjects of the big data and the main knowledge of library and information science, and put forward suggestions for data science courses suitable for library and information science. [Result/conclusion] The sequential topic network model can better fit the analysis of talent cultivation paradigm development in the big data subjects, and it may also be a way to research the topic of a subject. Finally, this paper puts forward some suggestions for the courses of data science which is oriented to library and information science, and there is certain reference value for the development of talent cultivation in library and information science.

Keywords: talent cultivation data science new liberal arts sequential topic network model